

Protein Purification Techniques Using the Intein Self-Cleaving Model

Undergraduate Honors Research Thesis
Submitted to the Engineering Honors Committee
119 Hitchcock Hall
College of Engineering
The Ohio State University
Columbus, Ohio 43210

Presented in partial fulfillment of the requirements for the designation of “Honors Research Distinction” for the degree of Bachelor of Science in Chemical and Biomolecular Engineering in the College of Engineering at The Ohio State University

By
Hannah C. Zierden
Bachelor of Science in Chemical and Biomolecular Engineering
The Ohio State University
2015

Thesis Committee:

Dr. David W. Wood, Advisor

Dr. Andre Palmer

Dr. Michael Paulaitis

Copyright by
Hannah Christine Zierden
2015

Abstract

A central step in the production of high purity recombinant proteins is their separation and purification. Recombinant proteins are expressed in host cells from which they are collected and purified. Purification is necessary to separate target recombinant proteins from the unwanted contents of the host cells in which they are grown. Common methods employ several chromatographic steps, which requires optimization, the use of expensive resins, and large time investments.

Another category of purification utilizes affinity tag sequences. Within this realm, affinity tags can be used in conjunction within the protein's naturally occurring, self-cleaving intein. Using the intein simplifies the purification into a one-step chromatography purification. In this method, an affinity tag, characterized by its ability to selectively bind to a specified resin, is fused to a target protein. This allows for the protein to be selectively separated from unwanted proteins and cell debris. Then, a pH shift induces the intein's self-cleaving capabilities, and the target protein can be separated from the resin and affinity tag.

Non-chromatographic affinity tags exist which eliminate the need for affinity resins. The elastin-like polypeptide (ELP) tag is one such sequence. The use of ELP tag with the self-cleaving intein, makes it possible for purifications to be done independent of an affinity column. Shifts in salt concentration and pH lead to successful purifications of the target protein. ELP has a large and repetitive protein sequence which requires large amounts of energy for synthesis. By shortening the length of ELP, expression can be increased by freeing some of that energy. In order to test this, and to determine the

optimal ELP tag length, which may be affected by size and solubility of the target protein, five different ELP tag sizes are studied.

Another tag, the maltose binding protein (MBP) affinity tag, is commonly used in chromatographic purification processes due to its ability to selectively bind to immobilized amylose. The MBP-tagged target protein binds to an amylose resin. Then, a pH shift causes the intein to undergo a cleaving reaction, allowing the target tag to be separated from the affinity sequence. In an effort to increase the economic feasibility and simplicity of the MBP purification, the amylose resin is replaced with a starch solution. The backbone of starch is primarily composed of amylose units. Starch, which contains negatively charged ionic groups, can be easily salted out of solution. By allowing starch, in solution, to bind to an MBP-tagged target protein, the target protein can be separated from other proteins and cell debris via a salting out method. The target protein is further purified via a pH shift in the purification buffer. Because starch and salts are relatively cheap, the success of this approach will lead to a new, feasible option for mass purification and production of proteins.

During the expression, proteins go through a folding process. In some cases, the recombinant proteins do not fold correctly, inactivating or altering the protein's functionality. This is problematic in cases when the protein will be used to develop vaccines, such as the third protein discussed here. Initial results showed that, during column purification, the protein becomes aggregated so it cannot be recovered from the column's affinity resin. In order to combat this problem, a Flag-Acidic-Target Tag (FATT) will be added to the protein using a polymerase chain reaction. FATT is made of three parts: the flag, the hyperacidic region and a target tag. The flag makes the tag

easily detected. The hyperacidic region expresses well in *E. coli*, is also highly charged, so can be purified in a single step using a standard anion exchange chromatography resin, and most importantly, has been shown to promote correct protein folding during expression. It aids in proper refolding of misfolded fusion partners containing disulfide bonds due to the structure of the hyperacidic region, which acts as a shield-like non-specific chaperone for the target protein during *in vitro* expression and refolding. The tagged protein will be purified using column purification where it will selectively adhere to the column and be separated at a high purity.

In all three experiments described here, protein yield will be determined using qualitative methods such as sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), and quantitative methods, including activity and Bradford assays.

Acknowledgements

I would like to thank Dr. David Wood for his guidance and support throughout the course of my undergraduate research career. I would also like to thank Tzu-Chiang Han, Samar Alsharawi, Dr. Michael Coolbaugh, and Dr. Richard Lease, members of the Wood Group, for their contributions to my knowledge, understanding, and work. I would like to thank Ashwin Lahiry, Angela Chen and my family for their moral support. I would like to especially thank Miriam Shakalli for her help, friendship, and shared love of Chinese food. Additionally, I would like to thank the Undergraduate Honors Committee within the College of Engineering at The Ohio State University for their financial support of my project. Finally, I would like to thank The Ohio State University and the William G. Lowrie Department of Chemical and Biomolecular Engineering for providing the resources necessary to become involved with undergraduate research and graduate with research distinction.

Vita

2011..... Cardington-Lincoln Local High School
2015..... B.S. Chemical and Biomolecular Engineering, The Ohio State University
2015 to Current Ph.D. Chemical Engineering, Johns Hopkins University

Fields of Study

Major Field: Chemical and Biomolecular Engineering

Table of Contents

Abstract	i
Acknowledgements	iv
Vita	v
List of Tables	vii
List of Figures	viii
I. Introduction and Background	1
II. Methodology	10
i. Chemicals and Reagents	10
ii. FATT Vector Construction	11
iii. Protein Expression	12
iv. ELP Purification	13
v. Starch Purification	15
vi. MBP Column Purification	15
vii. SDS-PAGE Analysis	16
III. Summary of Results	17
i. ELP Purification Results	17
ii. MBP Purification Results	19
iii. FATT Purification Results	23
IV. Conclusions	24
V. Personal Statement	26
VI. Bibliography	27

List of Tables

Table 1: Commonly used affinity resins and associated costs.....	4
Table 2: Protein weights in kilodaltons.	17

List of Figures

Figure 1: Mechanism of self-splicing intein ⁴	2
Figure 2: Mechanism of self-cleaving intein ⁴	2
Figure 3: Chromatography purification utilizing affinity tag in conjunction with self-cleaving intein ⁹	3
Figure 4: Aggregation of ELP tag as caused by temperature or salt concentration changes ¹²	5
Figure 5: Amino acid sequence of fused ELP tag, intein and target protein.	5
Figure 6: Structure of amylose, found in the backbone of starch, making starch a viable substitution for an amylose column ¹⁴	6
Figure 7: Lost functionality due to aggregated, or improperly folded protein.	7
Figure 8: FATT amino acid sequence and areas of interest. In this case, the cleavage site is the Factor Xa cleavage site. FATT can be designed with any desired cleavage site. The FLAG is also unnecessary for purification purposes ¹⁷	8
Figure 9: FATT in pUC57 ¹⁷	11
Figure 10: ELP purification scheme ¹²	14
Figure 11: SDS-PAGE for five ELP tag length purifications. Lanes: CL ₁ : clarified lysate for ELP tag length 1; SC ₁ : soluble contaminants for ELP tag length 1; W ₁ : wash for ELP tag length 1; P ₁ : Purified precursor for ELP tag length 1; CL ₂ : clarified lysate for ELP tag length 2; SC ₂ : soluble contaminants for ELP tag length 2; W ₂ : wash for ELP tag length 2; P ₂ : Purified precursor for ELP tag length 2; CL ₄ : clarified lysate for ELP tag length 4; SC ₄ : soluble contaminants for ELP tag length 4; W ₄ : wash for ELP tag length 4; P ₄ : Purified precursor for ELP tag length 4; CL ₇ : clarified lysate for ELP tag length 7; SC ₇ : soluble contaminants for ELP tag length 7; W ₇ : wash for ELP tag length 7; P ₇ : Purified precursor for ELP tag length 7; CL ₁₁ : clarified lysate for ELP tag length 11; SC ₁₁ : soluble contaminants for ELP tag length 11; W ₁₁ : wash for ELP tag length 11; P ₁₁ : Purified precursor for ELP tag length 11	18
Figure 12: MI-aFGF starch purification. Lanes: WL: whole lysate; CL: clarified lysate; SS: starch solution addition; SC: soluble contaminants; T0: time=zero sample; ON: time=16 hours sample; P: purified products; R: remnants; L: molecular weight ladder.	19

Figure 13: MI-GFP starch purification. Lanes: WL: whole lysate; CL: clarified lysate; SS: starch solution addition, .3%; SC: soluble contaminants, .3%; W: wash, .3%; T0: time zero sample, .3%; ON: time 16 hours sample, .3%; P: purified product, .3%; L: molecular weight ladder; SS _{0.4} : starch solution addition, .4%; SC _{0.4} : soluble contaminants, .4%; W _{0.4} : wash, .4%; T0 _{0.4} : time zero sample, .4%; ON _{0.4} : time 16 hours sample, .4%; P _{0.4} : purified product, .4%.	20
Figure 14: Fluorescence of GFP starch purification samples. Samples: WL: whole lysate; CL: clarified lysate; SS: starch solution; SC: soluble contaminants; W: wash; T0: time zero sample.	21
Figure 15: MI-GFP starch purification using suspended and autoclaved starch. Lanes: WL: whole lysate; CL: clarified lysate; SC: suspended starch soluble contaminants; W: suspended starch wash; T0: suspended starch time zero; ON: suspended starch time 16 hours; P: suspended starch purified product; R: suspended starch pellet remnants; SC _A : autoclaved starch soluble contaminants; W _A : autoclaved starch wash; T0 _A : autoclaved starch time zero; ON _A : autoclaved starch time 16 hours; P _A : autoclaved starch purified product; R _A : autoclaved starch pellet remnants.	21
Figure 16: Column purification of clarified lysate and soluble contaminants. Lanes: CL: clarified lysate; W8.5: pH 8.5 wash; W6.5: pH 6.5 wash; E1: pH 6.5 elution; E2: pH 8.5 elution; E3: second pH 8.5 elution; R: resin.	22
Figure 17: Initial purification results for Proteins C and M. Lanes: WL: whole lysate; CL: clarified lysate; SC: soluble contaminants; T0: time zero; ON: final time; P: purified product; R: resin.	23

I. Introduction and Background

A central step in the production of high purity recombinant proteins is their separation and purification. Purification is necessary to obtain target proteins from the host cells in which they are grown. Recombinant proteins can be used as antibodies, vaccines, enzymes and growth factors¹. Some common recombinant protein therapeutics include Amgen's Erythropoietin, Chiron's Hepatitis B vaccine and Genentech's tissue plasminogen activator¹. Without highly effective purification, proteins cannot be utilized in this broad range of areas. Furthermore, in order to mass-produce high purity recombinant proteins, the purification methods must be time and cost effective.

A promising tool used in the purification of several proteins is the self-cleaving intein. This biological macromolecule was developed from the naturally occurring self-splicing intein. The self-splicing intein, whose name comes from the fact that it is part of the *internal protein*, is embedded in-frame within a precursor protein sequence². Surrounded by exteins, this segment of the protein splices itself from the N- and C- terminals². This reaction begins with an N-O or N-S acyl rearrangement, followed by transesterification, a cyclization at the C-terminal and an acyl rearrangement³. The resulting protein segments include a mature protein product, made up of the extein segments, and individual segments of the excised inteins². Figure 1 shows an illustration of this splicing mechanism⁴.

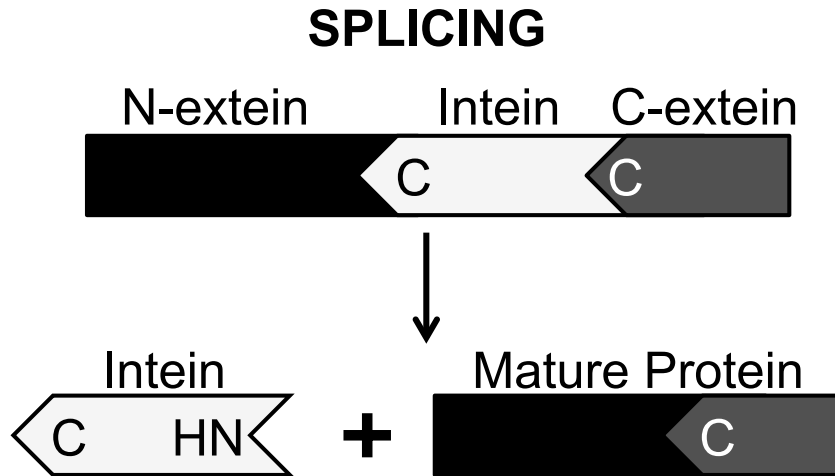


Figure 1: Mechanism of self-splicing intein⁴.

By mutating the naturally occurring N-terminal cysteine to alanine, the N-terminal cleavage function is disabled, thus eliminating the splicing nature of the intein^{4,5,6}. This results in a self-cleaving intein. This cleaving action has been optimized to occur under an induced by a pH shift from 8.0 to 6.0. An illustration of the self-cleaving intein is shown in Figure 2⁴.

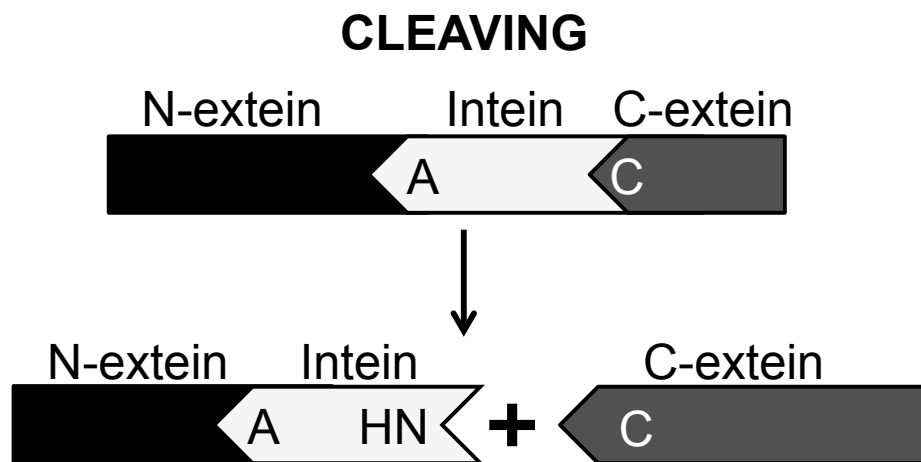


Figure 2: Mechanism of self-cleaving intein⁴.

The self-cleaving intein plays an important role for protein purification done with affinity

tags. Affinity tag sequences are characterized by their ability to selectively bind to a specified affinity resin^{7,8}. When fused to a target protein, an affinity tag allows for the selective separation of that target protein. By neglecting to remove the affinity tag, however, the functionality of the recombinant protein is affected. For example, the activity could be altered, or a therapeutic protein could become immunogenic⁷. While the affinity tag allows for the target protein to be separated from unwanted proteins and cell debris, the self-cleaving intein allows for the target protein to be separated from the affinity tag⁹. Figure 3 shows an illustration of an affinity tag purification utilizing the self-cleaving intein⁹.

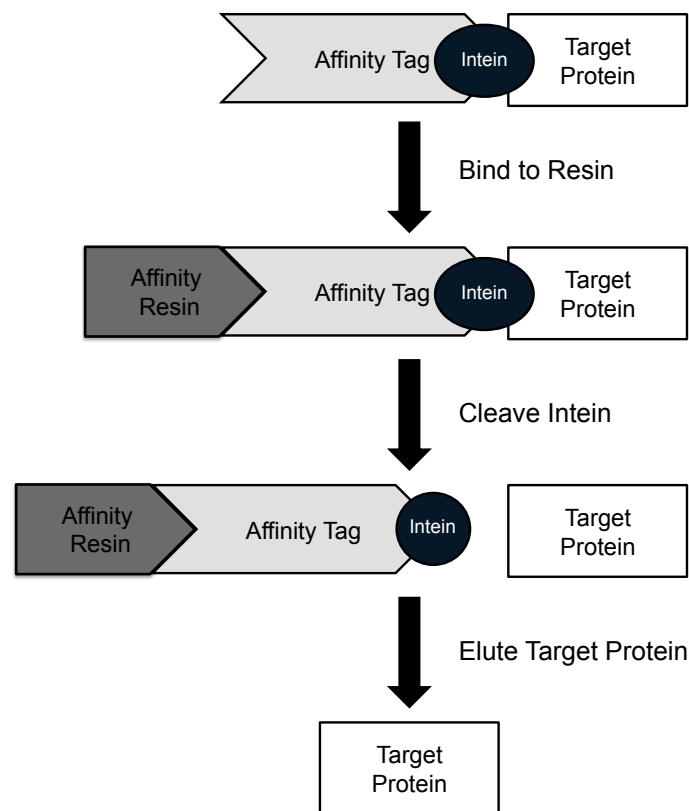


Figure 3: Chromatography purification utilizing affinity tag in conjunction with self-cleaving intein⁹.

Common protein purification techniques must be optimized for specific proteins, and

typically consist of several chromatographic steps. These steps involve expensive resins, large time investments and potentially low yields of the target protein¹⁰. Utilizing affinity tags simplifies the purification of nearly any target protein, but still requires optimization for the removal of the affinity tag—usually done using a protease enzyme¹¹. Protein purification is further simplified with the introduction of the intein. A previously multi-step chromatographic purification becomes a single-step purification when using an affinity tag in conjunction with the self-cleaving intein⁷. However, even in this simplified method, the cost of purification is high due to the cost of affinity resins. Table 1 gives the cost of some commonly used affinity resins⁷.

Table 1: Commonly used affinity resins and associated costs⁷.

Tag	Resin	Resin Cost (\$/g)	Capacity (mg/mL)	Regeneration Cycles	Manufacturer
Bulk Scale					
Chitin-binding domain	Chitin	180	2	5	NEB
His	Ni-NTA sepharose 6 Fast Flow	27	40	5	GE
Antibody Fc domain	MabSelect Xtra	3	42	100	GE
Laboratory Scale					
FLAG	M1 mAb	56,100	0.6	3	Sigma
MBP	Amylose	260	3	5	NEB

Non-chromatographic affinity tags exist which eliminate the need for affinity resins. These tags respond to changes in temperature and salt concentration⁷. The elastin-like polypeptide (ELP) tag is one such sequence. The use of ELP tag with the self-cleaving intein, makes it possible for purifications to be done independent of an affinity resin in a chromatography column. An increase in salt concentration causes the hydrophobic ELP tag to aggregate, forcing the tag-intein-target protein complex to precipitate out of solution¹². This phenomena is illustrated in Figure 4¹².

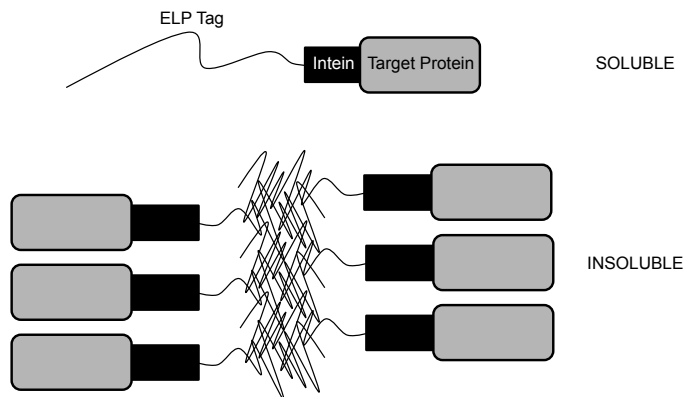


Figure 4: Aggregation of ELP tag as caused by temperature or salt concentration changes¹².

Using centrifugation and washing steps can ensure that the cell pellet contains only the ELP-tagged target protein. Then, a pH shift can cause the intein to cleave itself from the target protein, separating the target protein from the ELP-intein fusion. The tag will remain aggregate, allowing for the target protein to be separated and purified, independent of a chromatography column.

The protein sequence of the ELP tag is composed of multiple repeats of ten sets of the amino acid sequence VPGXG, wherein X represents any amino acid. This is shown in Figure 5¹².

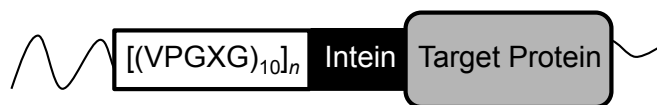


Figure 5: Amino acid sequence of fused ELP tag, intein and target protein.

During expression of the target protein, much of the cell's energy is devoted to synthesizing the ELP tag. If the number of repetitive ELP tag length sequences were to be diminished, the relieved energy could be devoted to the expression of the target protein, thus improving the yield of the purification. However, decreasing the length of

the ELP tag may affect the purification capabilities of the tag. In order to test the effect of tag length on protein purification, and to determine the optimal ELP tag length, five different ELP tag lengths are studied.

Another tag, the maltose binding protein (MBP) affinity tag, is commonly used in chromatographic purification processes due to its ability to selectively bind to immobilized amylose. The MBP-tagged target protein binds to an amylose resin, then a pH shift can cause the self-cleaving intein to undergo a cleaving reaction¹³. This allows the target tag to be separated from the affinity sequence. In an effort to increase the economic feasibility and simplicity of the MBP purification, the amylose resin is replaced with a starch solution. The backbone of starch is primarily composed of amylose units¹⁴. This molecule is shown in Figure 6.

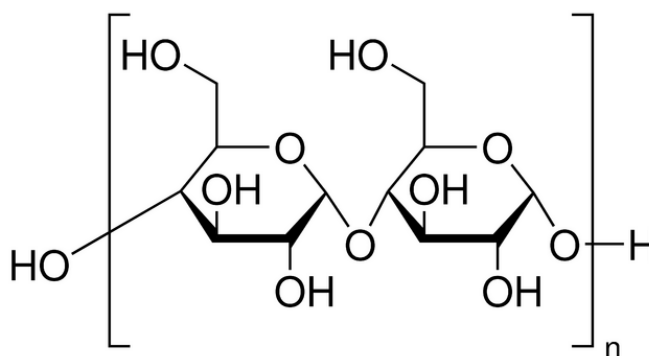


Figure 6: Structure of amylose, found in the backbone of starch, making starch a viable substitution for an amylose column¹⁴.

Starch, which contains negatively charged ionic groups, can be easily salted out of solution. Ionic repulsion in solution is weakened the presence of salt ions in solution¹⁵. By allowing starch to bind to an MBP-tagged target protein, the target protein can be

separated from other proteins and cell debris via a salting out method. The target protein is further purified through a pH shift in the purification buffer. The intein will cleave itself from the target protein, but remain attached to the MBP tag. The tag will remain a precipitate, and the target protein will be soluble in solution. Because starch and salts are relatively cheap, the success of this approach will lead to a new, feasible option for mass purification and production of proteins.

During the expression, proteins go through a folding process. In some cases, the recombinant proteins do not fold correctly, inactivating or altering the protein's functionality¹⁶. This is problematic in cases when the protein will be used to develop vaccines, such as the third protein discussed here. Initial purification results of the protein studied in this experiment, show that, during column purification, the protein becomes aggregated so it cannot be recovered from the column's affinity resin. This happening is depicted by the illustration in Figure 7.

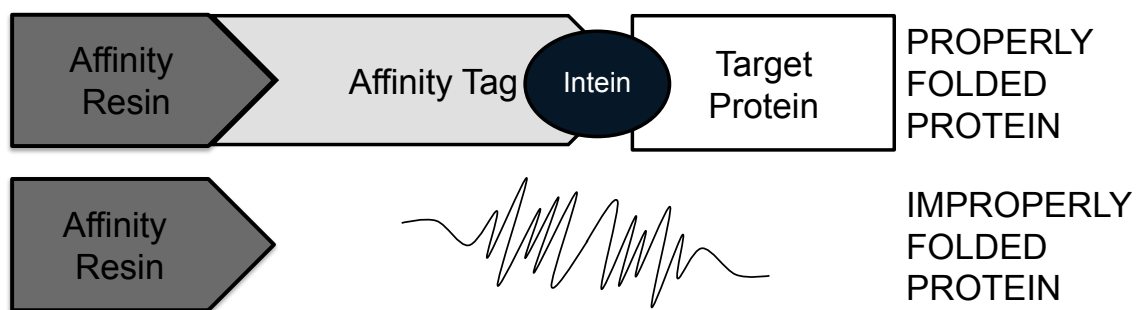


Figure 7: Lost functionality due to aggregated, or improperly folded protein.

In order to combat this problem, a Flag-Acidic-Target Tag (FATT) will be added to the protein using a polymerase chain reaction. FATT is made of three parts: the flag, the

hyperacidic region and a target tag¹⁷. The flag makes the tag easily detected. The hyperacidic region expresses well in *E. coli*, is also highly charged, so can be purified in a single step using a standard anion exchange chromatography resin, and most importantly, has been shown to promote correct protein folding during expression. It aids in proper refolding of misfolded fusion partners containing disulfide bonds due to the structure of the hyperacidic region¹⁷. This region acts as a shield-like non-specific chaperone for the target protein during *in vitro* expression and refolding¹⁶. Figure 8 shows the amino acid sequence of the FATT and its regions of interest.

<p>MDYKDDDDVE </p> <p>AEESDNVDSA</p> <p>DAEEDSDVW</p> <p>WGGADTDYAD</p> <p>GSEDKVVEVA</p> <p>EEEEVAEVEE</p> <p>EEADDDDEDE</p> <p>DGDEVVEEAE</p> <p>EPYEEATERT</p> <p>TSIATTTTTT</p> <p>TESVEEVYPG</p> <p>QVGYPGQVGY</p> <p>PGQVIEGRGI </p>	<p>The FLAG makes the tag easily detectable.</p> <p>The HYPERACIDIC REGION serves several purposes.</p> <ul style="list-style-type: none"> • Expresses well in <i>E. coli</i>. • Can be purified on an anion exchange chromatography resin. • Promotes proper protein folding during expression. • Aids in the refolding of misfolded fusion partners containing disulfide bonds • Acts as a shield-like non-specific chaperone for the target protein during <i>in vitro</i> expression and refolding. <p>TARGET TAG</p> <p>FACTOR XA cleavage site.</p>
---	--

Figure 8: FATT amino acid sequence and areas of interest. In this case, the cleavage site is the Factor Xa cleavage site. FATT can be designed with any desired cleavage site. The FLAG is also unnecessary for purification purposes¹⁷.

The tagged protein can be purified using anion exchange chromatography, where it selectively adheres to the column and is separated at a high purity. Then, a pH shift

forces the intein to cleave itself from the target protein, allowing the target protein to flow freely from the column.

II. Methodology

i. Chemicals and Reagents

For the ELP tag length experiments, stored stocks of the pET-E-I- β gal plasmid were grown up and expressed. The starch purification experiment utilized stored pET-M-I-X, where X was represented by the Acidic Fibroblast Growth Factor (aFGF), Enhanced Green Fluorescent Protein (eGFP) and Green Fluorescent Protein (GFP). In the FATT solubility purification, FATT was synthesized by GenScript © and cloned into the classified Proteins C and M, using the restriction enzymes NdeI and SacI from New England BioLabs ©.

The expression of all protein strains utilized Luria Bertani (LB)-agar plates with $100\mu\text{g}\cdot\text{mL}^{-1}$ ampicillin (amp), LB liquid media, Terrific Broth (TB) liquid media and Isopropyl- β -D-thiogalacto-pyranoside (IPTG). All growth media were sterilized in an autoclave at 121°C for 40 minutes. Ampicillin was sterilized via $0.2\mu\text{m}$ filtration.

The purification lysis buffer for the ELP and FATT purifications was composed of 10mM Tris-HCl at pH 8.6 and 2mM EDTA. The lysis buffer for the starch purification was composed of 200mM NaCl, 10mM Tris-HCl at pH 8.6 and 2mM EDTA. Buffering these solutions at pH 8.6 prevents unwanted intein cleavage, since the cleavage is induced around pH 6.5. Additionally, EDTA was used to prevent premature cleavage. Column purifications for the MBP tagged proteins required the use of an immobilized amylose resin from New England BioLabs ©. The column cleaving buffer for this

purification was composed of 100mM Tris-HCl at pH 6.5, 500 mM NaCl, 5% glycerol, 2mM EDTA and 1mM DDT. FATT purifications would utilize a Q-Sepharose Fast Flow from GE Healthcare.

Non-chromatographic methods used a variety of laboratory salts including: $(\text{NH}_4)_2\text{SO}_4$, CaCl_2 and NaCl. The starch purification of MBP-tagged proteins utilized starch from potato (Lot # SLBD6964V) obtained from Sigma Aldrich.

ii. FATT Vector Construction

The FATT sequence was synthesized by GenScript © into a pUC57 vector using NdeI and SacI. This sequence can be seen as Figure 9.

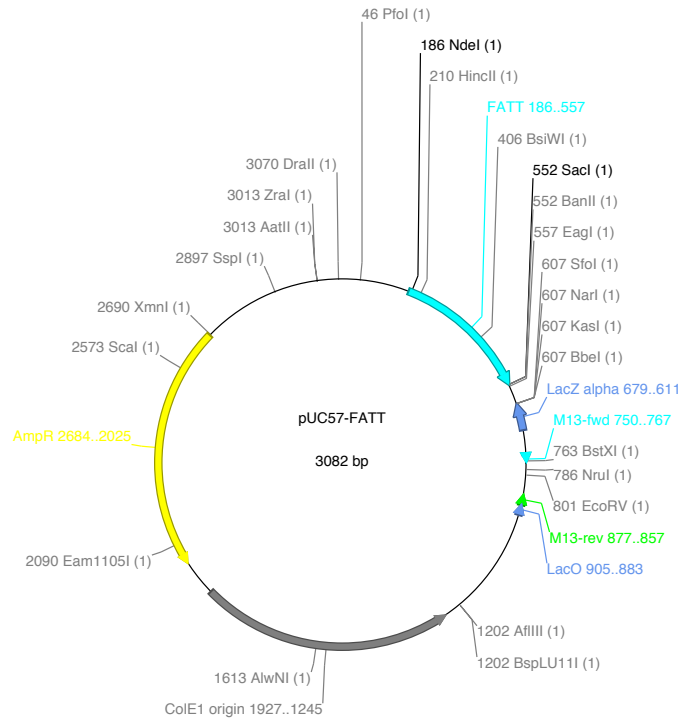


Figure 9: FATT in pUC57¹⁷.

The sample from GenScript © was double digested with NdeI and SacI restriction enzymes from NEB ©. The sample was run on an agarose gel and purified using QIAquick Gel. Proteins C and M were also digested with the same enzymes, isolating the vector through the same gel extraction method. The FATT insert was ligated with the backbone in a 6:1 ratio. The ligation was conducted at room temperature for 1 hour, and then transformed into DH5 α according to the Z-Competent *E. coli* Transformation Kit protocol by Zymo Research ©. The cells were plated onto an LB-agar plate with 100 $\mu\text{g}\cdot\text{mL}^{-1}$ amp. The clones were verified using digest checks and sequencing checks.

iii. Protein Expression

Protein expression allows for amplification of the target protein. To begin, stored plasmid stock was plated onto an LB-agar plate with 100 $\mu\text{g}\cdot\text{mL}^{-1}$ amp. This plate was allowed to grow for 14-16 hours at 37°C⁹. After this growth period, a single colony from the plate was chosen and transferred to 5mL of liquid LB media with 100 $\mu\text{g}\cdot\text{mL}^{-1}$ amp. This tube was left for 14-16 hours of growth in a water bath at 37°C and 200rpm⁹. The overnight was then diluted 1:100 into a baffled flask. This inoculation was supplemented with ampicillin, for a final concentration of 100 $\mu\text{g}\cdot\text{mL}^{-1}$. The flask was allowed to shake at 37°C and 180rpm for approximately three hours, or until the optical density of the cells reached between 0.8 and 1.0. Once reaching this value, the flask was cooled in a room temperature water bath for twenty minutes⁹. It was then induced with IPTG at a final concentration of 1mM. The induced culture was grown for 18 hours at 16°C and 180rpm. Cells were harvested in a centrifuge at 4000g for 10 minutes at 4°C. The collected cell pellets were resuspended in their respective lysis buffers and stored in a -80°C freezer for 12 hours⁹.

After this period of initial lysing, the cell suspensions were thawed on ice. Once completely thawed, the 1mL samples were sonicated, on ice, for 10 pulses of 20 seconds each at a setting of 0.6W. After sonication, 30 μ L of lysate was taken as the whole lysate sample⁹. This sample was used in SDS-PAGE and assay analysis.

The whole lysate was clarified using a centrifuge⁹. The sample was spun at 14,000g at 4 °C, and the supernatant was recovered. At this point, a 30 μ L clarified lysate sample was taken from the supernatant for later analysis. After separating the clarified lysate, specific purification protocols were followed.

iv. ELP Purification

Figure 10 shows the breakdown of the ELP purification¹².

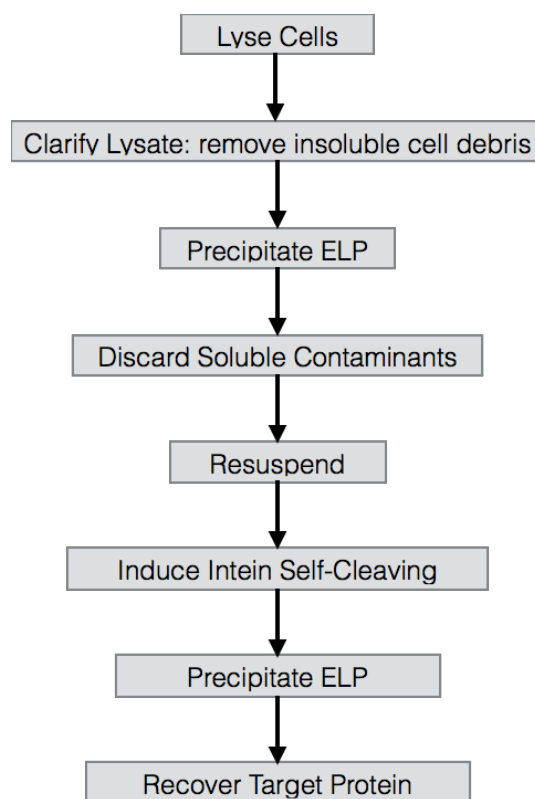


Figure 10: ELP purification scheme¹².

The clarified ELP cell lysate was mixed with $(\text{NH}_4)_2\text{SO}_4$ pH 8.0, for a final concentration of 0.4M. This mixture was gently mixed at room temperature for 30 minutes. The samples were centrifuged at 14,000g for 10 minutes at room temperature¹². The supernatant was taken as soluble contaminants. The pellet was washed twice with 0.4M $(\text{NH}_4)_2\text{SO}_4$ pH 8.0, and centrifuged at 14,000G for 10 minutes at room temperature¹². The supernatant was taken as the wash sample. The washed pellet was resuspended in 400 μ L of cleaving buffer (40mM Tris-HCl, pH 6.2). The constructs tested in this particular experiment utilized non-cleaving inteins. This was in an attempt to avoid confounding the ELP purification with the kinetics of intein cleavage. The resuspended pellet was taken as the purified precursor¹².

v. Starch Purification

For the starch precipitation purification, the supernatant, containing the target protein, was bound to a starch solution (200mM NaCl, 10mM Tris-HCl at pH 8.6, 2mM EDTA and 1% weight per volume starch) for a final concentration of 0.5% starch¹⁴. The mixture was gently agitated for one hour at room temperature. After an hour, 30µL of the starch solution was taken as a sample for further analysis. The starch-protein complex was then salted out of solution using the precipitation buffer (1M CaCl₂, 10mM Tris-HCl at pH 8.6, 2mM EDTA) After 10 minutes of precipitation, the pellet was be collected using a centrifuge spinning at 10,000g for 10 minutes¹⁴. The supernatant was taken as the soluble contaminants. The cell pellet was washed with the 400µL of precipitation buffer. The sample was centrifuged for 10 minutes at 10,000g, and the supernatant was taken as the wash sample. The cell pellet was then resuspended in maltose cleaving buffer (100mM Tris-HCl at pH 6.5, 500 mM NaCl, 5% glycerol, 2mM EDTA and 1mM DTT). A time-zero sample was taken after this addition. The mixture was be gently agitated for 16 hours at room temperature. After 16 hours, a final time sample was taken. Precipitation buffer was added to the sample for a final concentration of 500mM CaCl₂. This solution was gently agitated at room temperature for 10 minutes, and then centrifuged for 10 minutes at 10,000g. The supernatant was taken as the purified product¹⁴. The pellet was resuspended in lysis buffer for use in later analysis.

vi. MBP Column Purification

Before completing a column purification, the column first had to be prepared⁹. This was done by loading approximately 1mL of amylose resin into a standard 5mL gravity

column. The resin, which was stored in ethanol, had to be washed with 10 sample volumes of water, and 4 sample volumes of MBP lysis buffer (200mM NaCl, 10mM Tris-HCl at pH 8.6 and 2mM EDTA). The clarified lysate sample was diluted 1:5 with lysis buffer and loaded onto the column in 1mL intervals⁹. The flow-through of these loadings was collected as the initial load sample. After loading the entire sample, the column was washed with 12 volumes of lysis buffer. This flow-through was collected the pH 8.5 wash sample. The column was then washed with 12 volumes of cleaving buffer (100mM Tris-HCl at pH 6.5, 500mM NaCl, 5% glycerol, 2mM EDTA and 1mM DTT). The flow-through was collected as the pH 6.5 wash sample. The column was plugged and was incubated at room temperature for 18 hours. After 18 hours the column was washed with 1 volume of cleaving buffer. The flow-through was collected as the first elution sample. A second elution sample was collected from the flow-through of a single volume- lysis buffer wash⁹. After the elutions, a resin sample was taken by collecting 20μL of resin from the column, adding sample buffer (950μL Laemmli Sample Buffer from Bio-Rad ©, 50μL β-mercaptoethanol), heating the sample at 95°C for 10 minutes and centrifuging for 1 minute at 2,000g. The supernatant was taken as the resin sample⁹.

vii. SDS-PAGE Analysis

In order to check the results of the experiment sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was used. For each SDS-PAGE, the purification sample was mixed in a 1:1 ratio with sample buffer and heated at 95°C for 5 minutes. The samples were then run on a 5-12% polyacrylamide gel at 200V for 40 minutes. The gel was stained with Coomassie blue R-250 stain for viewing and analysis.

III. Summary of Results

SDS-PAGE was used to determine the results of the experiments. When running SDS-PAGE a molecular weight ladder was used to determine weights of the proteins in each sample. The weights for the proteins considered in these experiments can be found in Table 2.

Table 2: Protein weights in kilodaltons.

Protein	Weight (kDa)
E-I- β -gal	160.3
ELP-Intein	40-50
β -galactosidase	116.3
M-I-aFGF	59.5
MBP-Intein	42.5
aFGF	17.0
M-I-GFP	69.4
GFP	26.9
M-I-ProteinC	72.9
ProteinC	11.5
M-I-ProteinM	69.7
ProteinM	8.3

The weights are important for analysis in SDS-PAGE results.

i. ELP Purification Results

β -galactosidase purifications were done with five tag lengths of ELP: 1 repeat of the amino acid sequence, 2 repeats, 4 repeats, 7 repeats and 11 repeats. The results from these purifications can be seen in Figure 11. In this image, CL denotes the clarified lysate; SC denotes soluble contaminants; W denotes the wash sample; and P denotes the

purified precursor. Subscripts on these labels indicate the tag length at which the purification was completed. The numbers on the left of each gel indicate where the molecular weight marker protein sizes are.

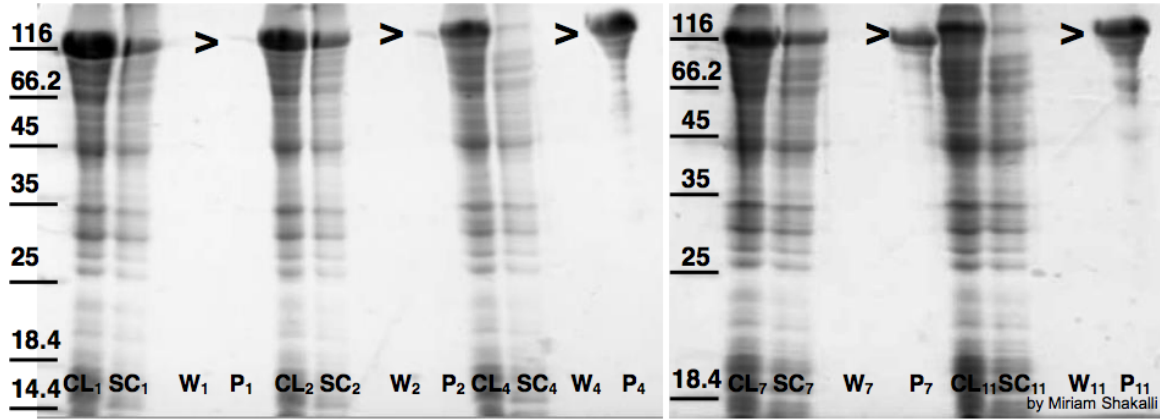


Figure 11: SDS-PAGE for five ELP tag length purifications. Lanes: CL₁: clarified lysate for ELP tag length 1; SC₁: soluble contaminants for ELP tag length 1; W₁: wash for ELP tag length 1; P₁: Purified precursor for ELP tag length 1; CL₂: clarified lysate for ELP tag length 2; SC₂: soluble contaminants for ELP tag length 2; W₂: wash for ELP tag length 2; P₂: Purified precursor for ELP tag length 2; CL₄: clarified lysate for ELP tag length 4; SC₄: soluble contaminants for ELP tag length 4; W₄: wash for ELP tag length 4; P₄: Purified precursor for ELP tag length 4; CL₇: clarified lysate for ELP tag length 7; SC₇: soluble contaminants for ELP tag length 7; W₇: wash for ELP tag length 7; P₇: Purified precursor for ELP tag length 7; CL₁₁: clarified lysate for ELP tag length 11; SC₁₁: soluble contaminants for ELP tag length 11; W₁₁: wash for ELP tag length 11; P₁₁: Purified precursor for ELP tag length 11.

In Figure 11, the tag length shows an effect on the recovery of the purified product. For ELP tag length 1, there appears a large “potato” band where β -galactosidase would be expected to show. However, very little of the protein is recovered in the purified precursor. As the tag length increases, more of the “potato” band from the clarified lysate lane is recovered in the purified precursor. However, in the ELP tag length 11 purification, the clarified lysate “potato” band is smaller than in any of the other purifications. This suggests that there is an equilibrium to be found between energy for expression and purification capacity of the tag. For the β -galactosidase purification, that

equilibrium lies at ELP tag length 4. In this purification, nearly all of the expressed protein is recovered, and the originally expressed protein shows good yield.

ii. MBP Purification Results

For the MBP purifications, two proteins were used. MI-aFGF was first used because of the clean cleaving that it has shown in past purifications. MI-GFP was used because of the green fluorescence that it shows when expressed. This characteristic allowed for the protein to be tracked throughout the purification.

Figure 12 shows an MI-aFGF purification, using 0.5% final starch concentration and 1M CaCl₂ as the precipitating agent.

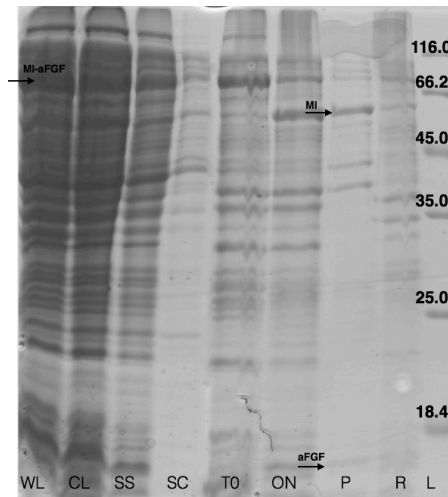


Figure 12: MI-aFGF starch purification. Lanes: WL: whole lysate; CL: clarified lysate; SS: starch solution addition; SC: soluble contaminants; T0: time=zero sample; ON: time=16 hours sample; P: purified products; R: remnants; L: molecular weight ladder.

In Figure 12, the purified product shows low recovery levels of aFGF. While some of the proteins are lost in the soluble contaminants, the T0 and ON samples show that there was low selectivity for starch precipitation. However, there is evident cleaving during the 16

hours of pH 6.5 agitation. Experiments were done comparing different precipitation buffers, varying amounts of starch in the starch solution and several temperatures for binding the protein to the starch. None of these experiments yielded better results. In an attempt to better see what was happening with the protein, purifications were done using MI-GFP. This strain exhibits fluorescence in expressed proteins, so the protein can be more easily tracked. Figure 13 shows the results of two MI-GFP purifications.

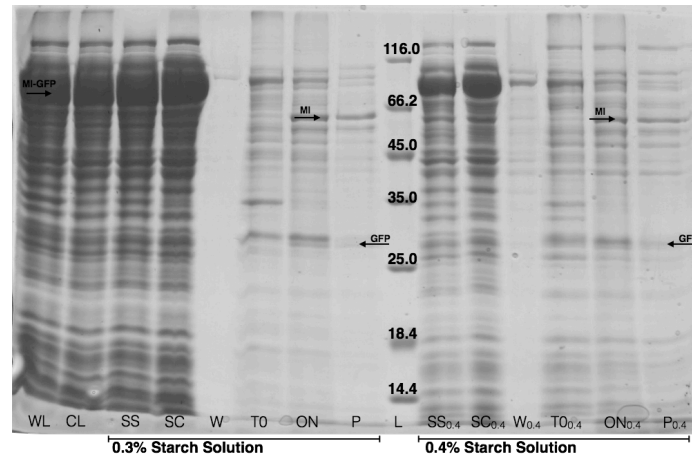


Figure 13: MI-GFP starch purification. Lanes: WL: whole lysate; CL: clarified lysate; SS: starch solution addition, .3%; SC: soluble contaminants, .3%; W: wash, .3%; T0: time zero sample, .3%; ON: time 16 hours sample, .3%; P: purified product, .3%; L: molecular weight ladder; SS_{0.4}: starch solution addition, .4%; SC_{0.4}: soluble contaminants, .4%; W_{0.4}: wash, .4%; T0_{0.4}: time zero sample, .4%; ON_{0.4}: time 16 hours sample, .4%; P_{0.4}: purified product, .4%.

As with the MI-aFGF results, the two different starch solutions yielded poor results for the purification of GFP. In this purification, nearly all of the protein was lost with the soluble contaminants. Samples for the 0.4% starch solution were examined under UV light, to better show their fluorescence. This is shown in Figure 14.

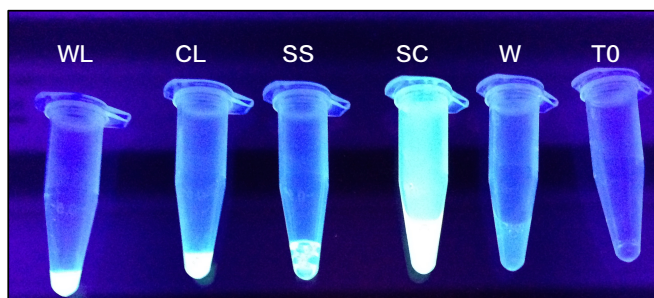


Figure 14: Fluorescence of GFP starch purification samples. Samples: WL: whole lysate; CL: clarified lysate; SS: starch solution; SC: soluble contaminants; W: wash; T0: time zero sample.

Figure 14 evidently shows that the target GFP is lost in the soluble contaminants. In an attempt to remedy this, a new starch solution was made. In this starch solution, the starch was autoclaved at 121°C for 40 minutes. This caused the starch granules to burst in solution, releasing smaller amylose molecules. A purification with this autoclaved starch solution was done and compared to the previously used starch solution. This comparison is shown in Figure 15.

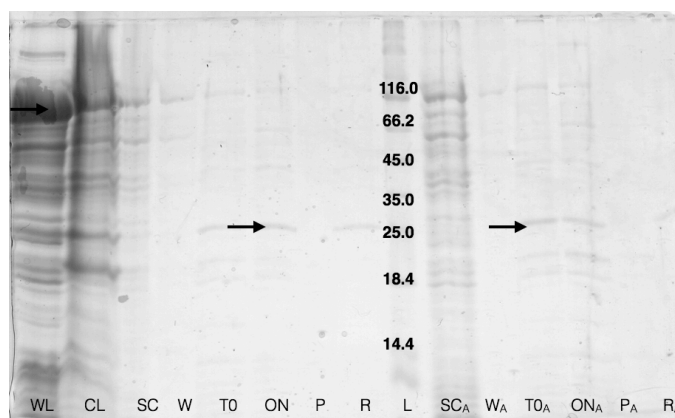


Figure 15: MI-GFP starch purification using suspended and autoclaved starch. Lanes: WL: whole lysate; CL: clarified lysate; SC: suspended starch soluble contaminants; W: suspended starch wash; T0: suspended starch time zero; ON: suspended starch time 16 hours; P: suspended starch purified product; R: suspended starch pellet remnants; SC_A: autoclaved starch soluble contaminants; W_A: autoclaved starch wash; T0_A: autoclaved starch time zero; ON_A: autoclaved starch time 16 hours; P_A: autoclaved starch purified product; R_A: autoclaved starch pellet remnants.

The comparison of starch solutions shown in Figure 16 showed no significant improvements between the suspended and autoclaved starch solutions. In an investigation of whether or not the starch was binding to the MBP, two column purifications were done. The first purification was done using a clarified lysate sample of MI-GFP. The second was done using the soluble contaminants from an autoclaved starch precipitation purification. This tested to see if the MBP was indeed bound to starch. The results of this experiment are shown in Figure 16.

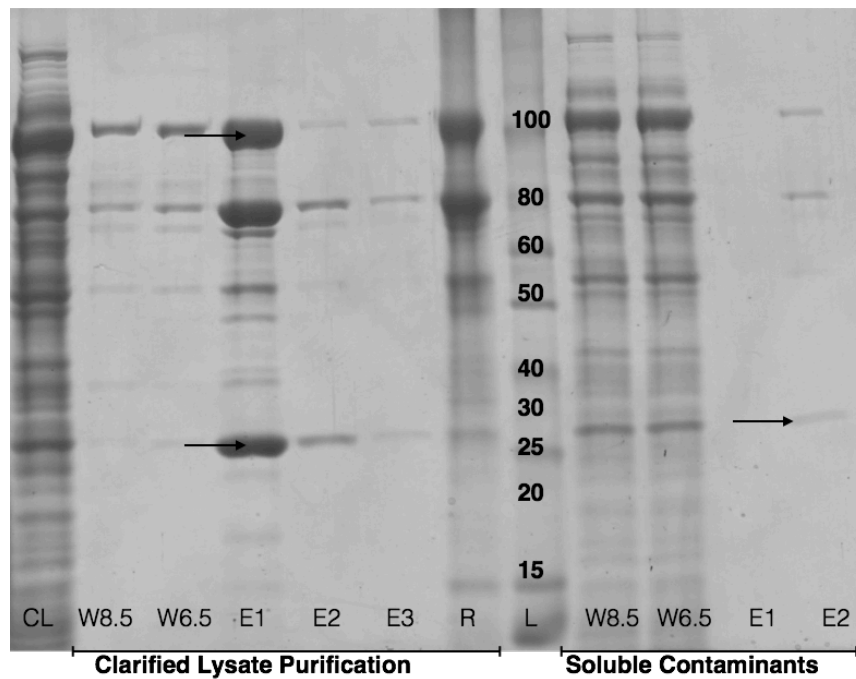


Figure 16: Column purification of clarified lysate and soluble contaminants. Lanes: CL: clarified lysate; W8.5: pH 8.5 wash; W6.5: pH 6.5 wash; E1: pH 6.5 elution; E2: pH 8.5 elution; E3: second pH 8.5 elution; R: resin.

Figure 16 shows a good column purification of GFP. The soluble contaminants purification gives indication that the starch is binding to the MBP, but that it is not precipitating out of solution. The wash lanes of soluble contaminants contain the MI-GFP band. If the amylose molecules from starch were not occupying the binding domain

of the maltose, the purification would have resembled the clarified lysate column purification. Rather, the soluble contaminants lost nearly all of the protein in the wash steps. No further purifications were done using starch and precipitating it out of solution.

iii. FATT Purification Results

An initial purification of Proteins C and M was completed using MBP. Figure 17 shows the initial SDS-PAGE results.

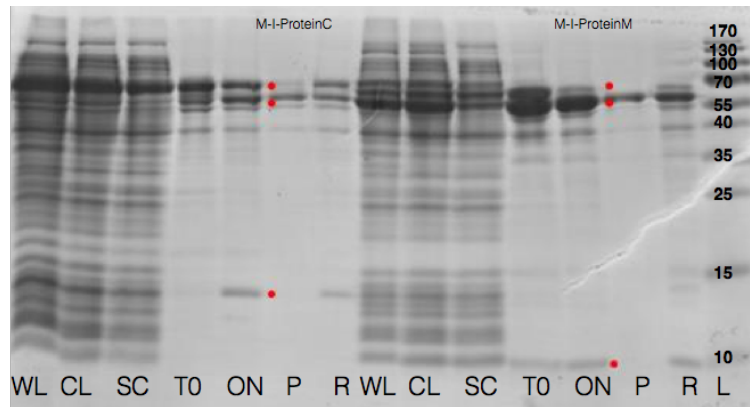


Figure 17: Initial purification results for Proteins C and M. Lanes: WL: whole lysate; CL: clarified lysate; SC: soluble contaminants; T0: time zero; ON: final time; P: purified product; R: resin.

In this purification, there is very little recovery of the target proteins. It is hypothesized that this is a result of protein insolubility due to improper folding. The FATT purification is expected to aid in this issue, but was unable to be completed. With more time, the cloning of FATT into Protein C and M will be completed and a purification will be done using an anion exchange column.

IV. Conclusions and Future Work

A promising tool in protein purification techniques is the self-cleaving intein. When used in conjunction with affinity tag sequences, this naturally occurring macromolecule leads to single-step purifications⁷. Several tags in existence are the elastin-like polypeptide (ELP), the maltose binding protein (MBP) and the flag acidic target tag (FATT). These tags aid in the specific separation of target proteins due to their specific separation characteristics⁸. The ELP tag can be forced into aggregation with shifts in salt concentration¹². MBP selectively binds to immobilized amylose, making it easily purified on an amylose affinity column⁹. FATT is purified on an anion exchange column¹⁷. By fusing any one of these tags to a target protein sequence, that target protein becomes easily separated from other proteins and cell debris. Use of the self-cleaving intein allows for tag removal under a pH shift. The tags were used to purify proteins in several experiments, making up the bulk of this work.

ELP tags are composed of repeating sequences of a 5-member amino acid sequence. Standard ELP tags include 110 repeats of this sequence. By decreasing the number of repeats, less energy is required for expression. This increases the yield of target protein in the cell lysate. Results showed that too few repeats diminishes the ELP's power to precipitate and function as a tag. However, as expected, too many repeats decreases the protein's expression. It can be concluded, that for β -galactosidase, an ELP tag length of 4 is ideal for purification purposes.

MBP is a tag that binds selectively to immobilized amylose¹⁴. The backbone of starch is primarily composed of amylose. By replacing an amylose affinity column with a starch solution, the target protein can be precipitated using a salting out approach¹⁵. This method improves the cost and efficiency of the purification. Results from this experiment show that the MBP tagged target proteins bind to starch, but are not salted out of solution. Future work could be done to investigate better salt precipitation tools. This would improve the purification and give rise to a new, inexpensive, time-efficient and feasible method of purification.

The third tag investigated in this study, FATT, aids in the proper folding and refolding of proteins^{16,17}. Its hyperacidic region acts as a shield to the protein in vitro, allowing it to fold properly. Because of frame shifts in cloning, the FATT tag purification has yet to be successful. Future work will include expression of FATT-tagged proteins, and their purification on anion exchange columns.

V. Personal Statement

During my four years at The Ohio State University, I have been provided resources and given opportunities to develop my scientific interests. I began my undergraduate research experience with Dr. David Wood in the William G. Lowrie Department of Chemical and Biomolecular Engineering. Over the past three years I have gained experience and knowledge in the field of protein purification, specifically in regards to biopharmaceuticals. I played a role in optimizing the purification method for a protein marketed by a French vaccine company. I completed many experiments aimed at investigating the effect of an elastin-like polypeptide (ELP) tag on purification results and I have transitioned into a mentor for new undergraduate researchers in the laboratory. I have worked on my own projects regarding the purification of maltose binding protein (MBP) tagged target proteins using starch and the affect that the flag-acidic target tag (FATT) has on the folding and refolding of proteins during expression. After my sophomore year, I was selected to participate in Germany's DAAD RISE program, where I further investigated protein purification as I studied the thermodynamics of protein adsorption. This work provided useful insight into the chromatographic methods used to purify and separate high purity proteins. These experiences, in addition to my coursework and efforts towards completing my honors thesis will greatly prepare me for my pursuit of a PhD. in Chemical Engineering at Johns Hopkins University starting in the fall of 2015.

VI. Bibliography

1. Koths, K. (1995). Recombinant proteins for medical use: The attractions and challenges. *Current Opinion in Biotechnology*, 6, 681-687.
2. Perler, F., Davis, E., Dean, G., Gimble, F., Jack, W., Neff, N., . . . Belfort, M. (1994). Protein splicing elements: Inteins and exteins — a definition of terms and recommended nomenclature. *Nucleic Acids Research*, 22(7), 1125-1127.
3. Chong, S., Mersha, F., Comb, D., Scott, M., Landry, D., Vence, L., . . . Xu, M. (1997). Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene*, 192, 271-281.
4. Wood, D., Wu, W., Belfort, G., Derbyshire, V., & Belfort, M. (1999). A genetic system yields self-cleaving inteins for bioseparations. *Nature Biotechnology*, 17, 889-892.
5. Shemella, P., Pereira, B., Zhang, Y., Roey, P., Belfort, G., Garde, S., & Nayak, S. (2007). Mechanism for Intein C-Terminal Cleavage: A Proposal from Quantum Mechanical Calculations. *Biophysical Journal*, 92, 847-853.
6. Wu, W., Wood, D., Belfort, G., Derbyshire, V., & Belfort, M. (2002). Intein-mediated purification of cytotoxic endonuclease I-TevI by insertional inactivation and pH-controllable splicing. *Nucleic Acids Research*, 30(22), 4864-4871.
7. Fong, B., Wu, W., & Wood, D. (2010). The potential role of self-cleaving purification tags in commercial-scale processes. *Trends in Biotechnology*, 28(5), 272-279.
8. Banki, M., & Wood, D. (2005). Inteins and affinity resin substitutes for protein purification and scale up. *Microbial Cell Factories*, 4(32).
9. Wood, D., Derbyshire, V., Wu, W., Chartrain, M., Belfort, M., & Belfort, G. (2000). Optimized Single-Step Affinity Purification with a Self-Cleaving Intein Applied to Human Acidic Fibroblast Growth Factor. *Biotechnology Progress*, 16, 1055-1063.
10. Banki, M., Gerngross, T., & Wood, D. (2005). Novel and economical purification of recombinant proteins: Intein-mediated protein purification using in vivo polyhydroxybutyrate (PHB) matrix association. *Protein Science*, 14, 1387-1395.

11. Arnau, J., Lauritzen, C., Petersen, G., & Pedersen, J. (2005). Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expression and Purification*, 48, 1-13.
12. Banki, M., Feng, L., & Wood, D. (2005). Simple bioseparations using self-cleaving elastin-like polypeptide tags. *Nature Methods*, 2(9), 659-662.
13. Young, C., Britton, Z., & Robinson, A. (2012). Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. *Biotechnology Journal*, 7, 620-634.
14. Raghava, S., Aquil, S., Bhattacharyya, S., Varadarajan, R., & Gupta, M. (2008). Strategy for purifying maltose binding protein fusion proteins by affinity precipitation. *Journal of Chromatography A*, 1194, 90-95.
15. Zhou, H., Wang, C., Shi, L., Chang, T., Yang, H., & Cui, M. (2014). Effects of salts on physicochemical, microstructural and thermal properties of potato starch. *Food Chemistry*, 156, 137-143.
16. Wood, D. (2014). New Trends and Affinity Tag Designs for Recombinant Protein Purification. *Current Opinion in Structural Biology*, 26, 54-61.
17. Sangawa, T., Tabata, S., Suzuki, K., Saheki, Y., Tanaka, K., & Takagi, J. (2013). A multipurpose fusion tag derived from an unstructured and hyperacidic region of the amyloid precursor protein. *Protein Science*, 22, 840-850.